



TU Clausthal

Fakultät für Energie- und Wirtschaftswissenschaften

Studiengang Wirtschaftsingenieurwesen

Statistische Analyse und Bewertung der Polymer-Simulation mcPolymer

Bachelorarbeit

zur Erlangung des akademischen Grades
Bachelor of Science (B.Sc.)

Marc Sauter

Matr.-Nr. 448280

Erstgutachter:	Prof. Dr. Jan Gertheiss Institut für Angewandte Stochastik und Operations Research
Zweitgutachterin:	Prof. Dr. Sabine Beuermann Institut für Technische Chemie

Inhaltsverzeichnis

1	Einleitung	4
2	Grundlagen	5
2.1	Radikalische Polymerisation	5
2.1.1	Reaktionsverlauf	6
2.1.2	Reaktionsführung	9
2.1.3	Charakterisierung von Polymerisaten	10
2.2	Simulation der radikalischen Polymerisation	12
2.2.1	Stochastisches Modell	12
2.2.2	Aufbau der Primärdaten	14
2.3	Statistische Methoden	16
3	Explorative Datenanalyse	25
3.1	Auswertung einer einzelnen Simulation	25
3.2	Auswertung innerhalb der Settings	27
3.3	Auswertung zwischen den Settings	29
4	Statistische Verfahren	35
4.1	Konfidenzintervalle für den empirischen Mittelwert	35
4.2	Zwei-Stichproben-t-Tests	37
4.3	Varianz-Tests	38
5	Zusammenfassung	39
A	Anhang	40
A.1	Datenbasis	40
A.2	Datenauswertung	40
A.3	Dokumentation	40

1 Einleitung

Simulationen von Polyreaktionen können einen Beitrag leisten die Mikrostruktur von Polymerisaten besser zu verstehen. Stochastische Modelle sind insbesondere für die Materialforschung das Mittel der Wahl. Ein weiteres Anwendungsfeld stellt die Vorhersage von Produktqualitäten in der Prozessindustrie dar.

Ziel der Arbeit ist die statistische Bewertung eines Datensatzes einer radikalischen Polymerisation, welcher mithilfe der Software *mcPolymer* unter Verwendung von Monte-Carlo-Methoden erzeugt wurde. Ein wesentlicher Erfolgsfaktor bei der Anwendung dieser Simulationen ist die Wahl einer geeigneten Anzahl an zu simulierenden Molekülen.

Zur Klärung der folgenden Forschungsfragen werden insgesamt 1000 Simulationen mit $1 * 10^{11}$ Molekülen sowie 2673 Simulationen mit $5 * 10^{10}$ Molekülen ausgewertet.

Es soll den Fragen nachgegangen werden wie sich die drei Zielgrößen Zahlenmittel der molaren Masse \overline{M}_n , Gewichtsmittel der molaren Masse \overline{M}_w und Umsatz des Monomers X_t (1) innerhalb einer Simulation über die Zeit, (2) innerhalb eines Settings von $1 * 10^{11}$ bzw. $5 * 10^{10}$ Molekülen sowie (3) zwischen den beiden Settings verhalten.

Im letzten Kapitel werden die wichtigsten Erkenntnisse zusammengefasst.

2 Grundlagen

Bei der Polymerisation erfolgt eine Verbindung von niedermolekularen Monomeren zu Makromolekülen [1, S. 3]. Grundsätzlich wird nach den beiden Reaktionsarten Kettenwachstumsreaktion und Stufenwachstumsreaktion unterschieden. Diese gliedern sich jeweils in weitere Synthesarten auf [1, S. 15].

Einschränkend wird in dieser Arbeit nur die radikalische Homopolymerisation betrachtet. Nach erfolgter Initiierungsreaktion bilden sich Polymere mit verzweigten Strukturen aus.

Zu Beginn der Bearbeitung stehen Simulationsergebnisse der Polymerisation von Butylacrylat zur Verfügung. Die Simulationen wurden am Institut für Technische Chemie an der Technischen Universität Clausthal mit dem Monte-Carlo Simulator *mcPolymer* durchgeführt und die Ergebnisse für die statistische Analyse bereitgestellt. Das Reaktionsmodell der Homopolymerisation von Butylacrylat wurde von Drache et al. publiziert [2]. Exemplarisch wurde das Beispiel $c_{BuA,0} = 1.5 \frac{\text{mol}}{\text{L}}$, 60°C ausgewählt.

Im Anschluss an das Grundlagenkapitel soll dieser Datensatz ausgewertet werden.

2.1 Radikalische Polymerisation

Zunächst wird der prinzipielle Verlauf der radikalischen Polymerisation und die zu erwartende Struktur der Polymere qualitativ erläutert.

Es folgen grundlegende Anmerkungen zur Durchführung der Reaktion.

Auf dieser Basis werden alle weiteren Zusammenhänge erarbeitet, die für die empirische Auswertung der Simulationsergebnisse notwendig sind.

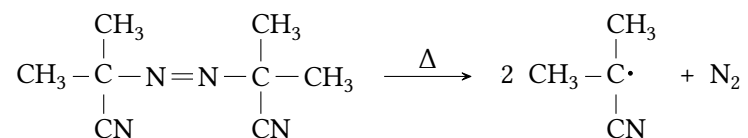
2.1.1 Reaktionsverlauf

Die radikalische Polymerisation unterteilt sich in die vier Schritte Initiierung, Start, Wachstum und Abbruch. Auf diese erste prinzipielle Darstellung folgt eine Erläuterung der zusätzlich ablaufenden Transferreaktionen.

Initiierung

Die Initiierung der Reaktion kann auf verschiedene Weisen ausgelöst werden. Ziel ist immer eine Bildung von freien Radikalen [1, S. 62 ff.]. Diese besitzen ein ungepaartes Elektron und zeichnen sich damit durch eine hohe Reaktivität aus [3, S. 863].

In der auszuwertenden Datenbasis wird explizit der Initiator Azoisobutyronitril (AIBN) verwendet. Ausgehend von Raumtemperatur zerfällt dieser bei leichter Wärmezugabe in zwei Cyanopropylradikale sowie ein Stickstoffmolekül [1, S. 62]. Der elementare Stickstoff entweicht dabei als Gas und ohne Beteiligung an weiteren Reaktionen.



In Relation sind nur sehr kleine Mengen eines Initiators notwendig.

Start

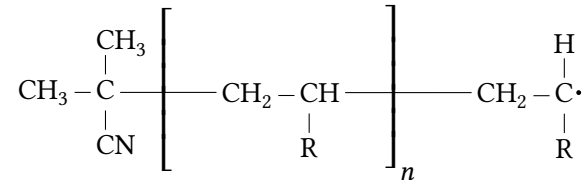
Neben dem Initiator müssen im Reaktionsgemisch auch die zu polymerisierenden Monomere in ausreichender Anzahl vorliegen. Deren strukturelle Voraussetzung ist das Vorhandensein einer Doppelbindung [4, S. 64].

Unter Aufbrechen der Kohlenstoffdoppelbindung $\text{C}=\text{C}$ des Monomers lagert sich ein freies Radikal R^\cdot an ein Monomer M an. Die Radikalfunktion geht dabei nicht verloren sondern überträgt sich auf das freie Ende des Monomers [4, S. 73].

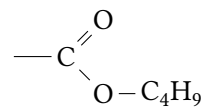
Die Startreaktion der radikalischen Polymerisation ist per Definition mit der Anlageung des ersten Monomers abgeschlossen.

Wachstum

Während der Wachstumsreaktion addieren sich an das zuletzt angelagerte Monomer weitere Monomere jeweils unter Erhalt der Radikalfunktion [1, S. 66]. Es entsteht ein Polymer mit Polymerisationsgrad n und folgender Konstitution.



Für Butylacrylat steht der Rest R für



Abbruch

Der Abbruch der radikalischen Polymerisation kann auf zwei Arten erfolgen.

Bei der Kombination verbinden sich Radikale, es entsteht eine einzige Polymerkette.

Bei der Disproportionierung entstehen durch Abstrahierung eines Wasserstoffatoms jeweils eine ungesättigte und eine gesättigte Polymerkette [5, S. 415].

In der Regel liegen beide Abbruchmechanismen gleichzeitig vor. Thermodynamisch bevorzugt ist die Kombination von Radikalen, da hierfür eine geringere Aktivierungsenergie benötigt wird [4, S. 77]. Dieser Abbruchmechanismus geht immer mit einer Erhöhung des mittleren Polymerisationsgrades einher [4, S. 77].

Übertragung auf Polymere

Von dem Idealtyp der linearen Kettenbildung abweichend kommt es zusätzlich zu Übertragungsreaktionen auf Polymere [1, S. 76 ff.]. Es macht Sinn, sich vorab die zu erwartende Struktur vorzustellen, wie sie bisher beschrieben wurde.

Die Länge einer Polymerkette ist statistisch zufällig verteilt und hängt unter anderem von dem Abbruchmechanismus ab. Als generelles konstituierendes Merkmal sind die Kohlenstoffeinfachbindungen zu nennen. Diese ermöglichen eine nahezu freie Drehbarkeit im Raum [3, S. 699 ff.]. Die Strukturvielfalt für zwei Polymerketten mit konstantem Polymerisationsgrad n ist anhand Abbildung 2.1 angedeutet.

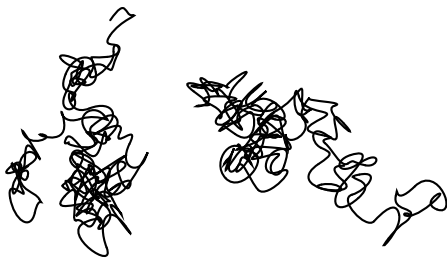


Abb. 2.1: Unverzweigte Polymerketten

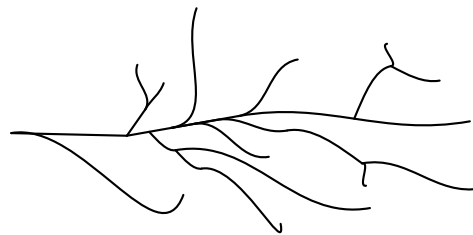


Abb. 2.2: Ausschnitt einer verzweigten Kette

Mit dem Mechanismus der Übertragung ist gemeint, dass von einer wachsenden Kette die Radikalfunktion auf eine andere Stelle übergeht. Dies geschieht unter Abstraktion eines Wasserstoffatoms in Richtung des einst vorhandenen Radikals [5, S. 415].

Die Übertragung der Radikalfunktion kann auf eine beliebige funktionelle Gruppe eines anderen Polymers erfolgen und führt zu Langkettenverzweigungen (LCB). Diese sind in Abbildung 2.2 angedeutet. Als lang kann man die Seitenzweige deshalb bezeichnen, weil sie eine gleich hohe Wachstumswahrscheinlichkeit wie das angenommene Polymerkettenende besitzen [4, S. 82].

An praktisch jeder Stelle eines Polymers kann aber auch eine Kurzkettenverzweigung (SCB) vorliegen. Der zugrundeliegende Mechanismus ist eine intramolekulare Übertragung der Radikalfunktion und wird als back-biting bezeichnet [4, S. 82]. Diese Verzweigungen entstehen immer nahe am Ende einer wachsenden Polymerkette.

Weitere Nebenreaktionen

Neben Polymeren können auch Monomere oder Lösungsmittelmoleküle von der Übertragung der Radikalfunktion in unterschiedlichem Maße betroffen sein. Weiterhin ist es möglich, durch den Einsatz von Kettenreglern (organische Verbindungen wie Thiole, die leicht ein Wasserstoffatom abgeben) den Polymerisationsgrad in der Tendenz zu erniedrigen.

Die Übertragung der Radikalfunktion auf Monomere ist der bekannten Wachstumsreaktion ähnlich. Das verwendete Lösungsmittel THF ist nur im geringen Maße von der Kettenübertragung betroffen, diese Reaktion ist im kinetischen Schema enthalten [2].

2.1.2 Reaktionsführung

Je nach angewandter Reaktionstechnik variiert der Ablauf der radikalischen Polymerisation. Im konkreten Anwendungsfall wird die Polymerisation homogen als Lösungspolymerisation simuliert. Der Reaktor hat ein konstantes Volumen und wird als ideal durchmischt angenommen. Dieser wird diskontinuierlich betrieben. Die Konzentration frei verfügbarer Monomere c_M hat also zu Beginn ihr Maximum und fällt stetig ab.

Zeitabhängigkeit

Die radikalische Polymerisation verläuft sehr schnell. Eine wachsende Kette lagert etwa ein Monomer pro Millisekunde an [5, S. 416].

Mit zunehmender Verknäulung der sich bildenden Polymere stellt sich eine Verlangsamung der Abbruchreaktion ein. Dies ist auf die verminderte Diffusionsfähigkeit der Moleküle zurückzuführen [3, S. 791 ff.]. Polymerisationen werden bei dem Beispiel der Polymerisation von Butylacrylat aus praktischen Gründen nach einer definierten Zeit abgebrochen [1, S. 70 f.], sodass noch unverbrauchte Monomere vorliegen.

Temperaturabhängigkeit

Die Höhe der Temperatur hat einen entscheidenden Einfluss auf den Reaktionsverlauf. Grundsätzlich handelt es sich bei der Polymerisation um eine exotherme Reaktion. Die Ceiling-Temperatur ist als obere Grenze zu verstehen, ab der keine Polymerisation mehr möglich ist [1, S. 79]. Ebenso findet die Polymerisation auch bei zu geringer Temperatur nicht statt, da dort der Initiator AIBN zu langsam zerfällt.

Initiatorabhängigkeit

Mit steigender Konzentration des Initiators c_{AIBN} steigt die Anzahl der gebildeten Radikale und damit auch die Anzahl der Ketten, die gestartet werden. Bleibt die Konzentration an verfügbaren Monomeren c_M gleichzeitig konstant, werden die gebildeten Ketten in der Folge kürzer sein.

Die Radikalausbeute f bezeichnet das Verhältnis der tatsächlich zur radikalischen Polymerisation verwendeten Radikale zur eingesetzten Gesamtanzahl [4, S. 70]. Auf Basis eines Literaturwertes wird dieser Wert mit 0.64 angenommen [2].

2.1.3 Charakterisierung von Polymerisaten

Typisch für Produkte der radikalischen Polymerisation ist der unterschiedliche Polymerisationsgrad der enthaltenden Polymere. Zur Charakterisierung von Polymerisaten hat sich das Aufstellen der Verteilungsfunktion der molaren Masse etabliert.

In der Praxis wird diese analytisch ermittelt. Dazu bedient man sich in der Regel der Gelpermeations-Chromatografie (GPC). Durch Ausnutzung des Größenausschlusses in Poren, d.h. der Trennung der Moleküle nach ihrer Größe, kann deren Verteilung indirekt bestimmt werden [6, S. 424 f.].

Ziel soll es nun sein, den Sinngehalt der Verteilungsfunktion der molaren Masse und die daraus abgeleiteten statistischen Maßzahlen zu verstehen.

Stoffmenge

Rechnungen mit der Masse $m(X)$ sind in der Chemie nicht zweckmäßig, da für eine gegebene Masse unterschiedliche Stoffe eine unterschiedliche absolute Teilchenanzahl $N(X)$ besitzen. Stöchiometrisch einfache Zahlenverhältnisse erreicht man mit der Stoffmenge $n(X)$ unter Zuhilfenahme der Avogadro-Konstante N_A .

$$[n(X)] = \frac{[N(X)]}{[N_A]} = 1 \text{ mol} \quad (2.1)$$

Die Menge der Teilchen, aus denen sich die in Abschnitt 2.1.1 erläuterten Moleküle zusammensetzen können, sind die chemischen Elemente Wasserstoff (H), Kohlenstoff (C), Stickstoff (N), und Sauerstoff (O).

Stoffmengenkonzentration

Die Stoffmengenkonzentration $c(X)$ ist ein Maß für die Stoffmenge in einem betrachteten Volumen $V(X)$. Das Volumen soll konstant sein.

$$[c(X)] = \frac{[n(X)]}{[V]} = 1 \frac{\text{mol}}{\text{m}^3} = 1 \times 10^{-3} \frac{\text{mol}}{\text{L}} \quad (2.2)$$

Üblicherweise erfolgt die Angabe auf Basis der Einheit Liter.

Molare Masse

Die molare Masse $M(X)$ ist eine für die Polymercharakterisierung zentrale Größe. Sie ist aus der Masse $m(X)$ eines Stoffes und der Stoffmenge $n(X)$ ableitbar.

$$[M(X)] = \frac{[m(X)]}{[n(X)]} = 1 \frac{\text{kg}}{\text{mol}} = 1 \times 10^3 \frac{\text{g}}{\text{mol}} \quad (2.3)$$

Üblicherweise erfolgt die Angabe auf Basis der Einheit Gramm. Für das Monomer Butylacrylat beträgt die molare Masse $M_{BuA} = 128.17 \frac{\text{g}}{\text{mol}}$.

Polymer

Es wird die weitere Vereinfachung getroffen, dass das Initiatorfragment bei der Charakterisierung eines Polymerisats hinsichtlich seiner Masse nicht berücksichtigt wird.

Für Homopolymere, Polymere die aus nur einer Monomerart aufgebaut sind wie es bei den zugrunde liegenden Simulationen der Fall ist, kann die molare Masse eines Polymers damit vereinfachend als Vielfaches der molaren Masse des Monomers errechnet werden.

Es gilt für den hier beabsichtigten Anwendungszweck die folgende Beziehung.

$$M(X) = n * M_{BuA} \quad (2.4)$$

Verteilungsfunktion

In der Einleitung des Abschnitts wurde bereits erwähnt, dass zur Charakterisierung eines Polymerisats häufig eine Verteilungsfunktion aufgestellt wird.

Im einfachsten Fall stellt man in einem kartesischen Koordinatensystem die Kettenlänge auf der horizontalen Achse dar, während die Anzahl der Polymere mit der jeweiligen Kettenlänge auf der vertikalen Achse aufgetragen werden.

Lagemaße

Ein Lagemaß ist ein numerischer Wert, der das Zentrum einer Verteilung beschreibt [7, S. 49]. Die beiden Lagemaße Zahlenmittel der molaren Masse \overline{M}_n und Gewichtsmittel der molaren Masse \overline{M}_w werden in der Prozessindustrie zur Kurzcharakterisierung eines Polymerisats genutzt. Diese sollen auch in der Datenauswertung Anwendung finden und bedürfen daher an dieser Stelle einer kurzen Ausführung.

Das Zahlenmittel \overline{M}_n errechnet sich in Anlehnung an [1, S. 9] wie folgt.

$$\overline{M}_n = \frac{\sum n * M(X)}{\sum n} \quad (2.5)$$

Das Gewichtsmittel \overline{M}_w errechnet sich in Anlehnung an [1, S. 9] wie folgt.

$$\overline{M}_w = \frac{\sum m * M(X)}{\sum m} = \frac{\sum n * (M(X))^2}{\sum n * M(X)} \quad (2.6)$$

Umsatz des Monomers

Der Umsatz des Monomers X_t ist ein Maß für den Reaktionsfortschritt. Er kann Werte im Intervall von 0 bis 1 annehmen und wächst mit der Zeit. Der Umsatz des Monomers ist wie folgt definiert.

$$X_t = \frac{n_0 - n_t}{n_0} \quad (2.7)$$

2.2 Simulation der radikalischen Polymerisation

Die Motivation bei der Simulation der radikalischen Polymerisation besteht darin Modellvorstellungen zu verifizieren.

Eine Simulation der radikalischen Polymerisation kann auf grundsätzlich zwei verschiedene Arten erfolgen. Rein deterministische Ansätze bieten den Vorteil, dass Ergebnisse sehr schnell berechnet werden können. Stochastische Modelle kommen mit der gleichen Parameteranzahl aus, deren Entwurf und die Berechnung von Ergebnissen ist aber in der Regel mit einem deutlichen Mehraufwand verbunden. Der Mehraufwand von Simulationen mit einem stochastischen Modell kann dafür mit einer höheren Informationstiefe belohnt sein.

Im Folgenden wird ein Überblick über das verwendete stochastische Modell gegeben. Schließlich werden die vorliegenden Primärdaten beschrieben.

2.2.1 Stochastisches Modell

Das für die Generierung der Datenbasis zugrunde liegende mathematische Modell geht auf eine Publikation von D.T. Gillespie zurück [8]. Es werden Monte-Carlo-Methoden verwendet, mit deren Hilfe Polymerketten als einzelne Objekte simuliert werden [9].

Die Verwendung von Monte-Carlo-Methoden folgt zwei Prinzipien: Erstens wird die nächste auszuführende chemische Reaktion durch Pseudo-Zufallszahlen bestimmt [10]. Zweitens werden die Ergebnisse stetig besser, wenn die Anzahl an simulierten Molekülen steigt [10].

Der Simulator ist im Ordner „master“ abgelegt und bedarf weiterer Erläuterungen.

Reaktorvolumen

Der Reaktor wird als ideal durchmischt angenommen und soll ein konstantes Volumen besitzen [9]. Die Simulation startet mit einer gegebenen Molekülanzahl.

Mit der Publikation der Software *mcPolymer* konnte bereits eine untere Grenze der Molekülanzahl bestimmt werden, die für korrekte Ergebnisse nicht unterschritten werden darf [9].

Die relevante Frage hier lautet, ob das mehr an Informationstiefe einen erhöhten Berechnungsaufwand rechtfertigt. Daraus leitet sich die Motivation für die statistische Untersuchung ab.

Die Simulationen werden mit $1 \cdot 10^{11}$ bzw. $5 \cdot 10^{10}$ Molekülen durchgeführt. Dies entspricht jeweils dem Kontrollvolumen der Monte-Carlo-Simulation.

Reaktionswahrscheinlichkeit

Die Partikel X_i der chemischen Spezies i sind in dem Reaktor vorhanden und haben jeweils eine gewisse Reaktionswahrscheinlichkeit [9]. Die Reaktionswahrscheinlichkeiten berechnen sich aus der Molekülanzahl der Spezies und den Reaktionsgeschwindigkeitskoeffizienten.

Mithilfe einer ersten Pseudo-Zufallszahl wird unter Berücksichtigung der Reaktionswahrscheinlichkeiten eine Reaktion ausgewählt.

Mit einer zweiten Pseudo-Zufallszahl wird aus einer Gesamtreaktionswahrscheinlichkeit das Zeitintervall τ für einen Monte-Carlo-Schritt bestimmt.

Von Pseudo-Zufallszahlen wird deshalb gesprochen, weil diese mit einem Algorithmus deterministisch berechnet werden und sich nur annähernd wie echte Zufallszahlen verhalten.

Weitere Simulationsparameter

Die Daten werden im Zeitintervall von 10 Sekunden erfasst. Die Protokollierung der Daten endet nach einer Simulationszeit von 3600 Sekunden. Die Reaktionstemperatur ist auf 60 °C festgelegt. Die Initiatorkonzentration nimmt den Wert $0.01 \frac{\text{mol}}{\text{L}}$ an. Desweiteren sind im Simulator die chemischen Spezies, die möglichen chemischen Reaktionen sowie die zugehörigen Reaktionsgeschwindigkeitskoeffizienten hinterlegt.

2.2.2 Aufbau der Primärdaten

Um die Primärdaten empirisch auswerten zu können ist es zweckmäßig zunächst einen Überblick über deren grundsätzlichen Aufbau zu gewinnen.

Die Daten sind in durchnummerierten Dateiodnern angelegt.

Für eine Molekülanzahl von $1 * 10^{11}$ (im Ordner 1e11) sind dies die Dateiodner sim1 bis sim1000, für eine Molekülanzahl von $5 * 10^{10}$ (im Ordner 5e10) sind dies die Dateiodner sim1 bis sim2673.

Die Simulationsergebnisse stellen zufällige Stichproben mit gleichen Reaktionsbedingungen dar und sind von einander unabhängig. Sie unterscheiden sich nicht in der Anzahl der Dateien und auch nicht in deren Bezeichnungen. Lediglich der Dateiinhalt kann variieren. Als Übersicht sind die Dateinamen mit einer Kurzbeschreibung in Tabelle 2.1 zusammengefasst.

Die Dateiinhalte werden nachfolgend detailliert beschrieben.

cAIBN.0.dat, cBuA.0.dat, cBuAX.0.dat

In diesen Dateien werden Konzentrations-Zeit-Verläufe erfasst. Die Spalten enthalten Werte für „time“ in Sekunden, „c“ in $\frac{\text{mol}}{\text{L}}$ sowie „molecules“ als absolute Anzahl an Molekülen. Entsprechend der Vorgabe im Simulator werden die Werte im Intervall von 10 Sekunden protokolliert.

Die erste Zeile bezeichnet den Spalteninhalt. Die zweite Zeile gibt die Konzentration und die Anzahl der Moleküle der chemischen Spezies zum Zeitpunkt 0 Sekunden wieder, bevor die erste chemische Reaktion abläuft. Dem Zeitintervall entsprechend enthält Zeile 362 Informationen über die Konzentration und die Anzahl der Moleküle der chemischen Spezies nach 3600 Sekunden. Die Protokollierung der Daten endet mit diesem Zeitpunkt.

Tab. 2.1: Aufbau eines Simulationsergebnisses

Konzentrations- Zeit-Verläufe	cAIBN.0.dat	Initiator
	cBuA.0.dat	Monomer
	cBuAX.0.dat	Monomer (Dummy)
Konzentrations- Zeit-Verläufe (detailliert)	cD.0.dat	
	cI#.0.dat	
	cLM.0.dat	
	cMCR#.0.dat	
	cMCRL#.0.dat	
	cSPR#.0.dat	
Kettenlängenverteilungen	D.120.0.cld	nach 120 Sekunden
	D.600.0.cld	nach 600 Sekunden
	D.1200.0.cld	nach 1200 Sekunden
	D.1800.0.cld	nach 1800 Sekunden
	D.2400.0.cld	nach 2400 Sekunden
	D.3000.0.cld	nach 3000 Sekunden
	D.3600.0.cld	nach 3600 Sekunden
Logdateien	log.dat	aktuelle Kopie der Logdatei aus dem Simulator „master“
	ktlog.dat	Schätzungen der Abbruchparameter für den Fall der Kombination

Die Dateien cAIBN.0.dat und cBuA.0.dat enthalten neben der ersten Zeitangabe ausschließlich Werte ungleich 0, da der Initiator AIBN und das Monomer Butylacrylat tatsächlich an der Reaktion beteiligt sind und zu Beginn der Reaktion vorliegen. Wie bereits in Abschnitt 2.1.2 angedeutet nehmen die Werte der Konzentration bzw. die absolute Teilchenanzahl stetig ab.

cD.0.dat, cI#.0.dat, cLM.0.dat, cMCR#.0.dat, cMCRL#.0.dat, cSPR#.0.dat

Der Aufbau der Dateien ist den vorhergehenden sehr ähnlich. Es existieren wieder drei Spalten mit den Bezeichnungen „time“ in Sekunden, „c“ in $\frac{\text{mol}}{\text{L}}$ sowie „molecules“ als absolute Anzahl an Molekülen. Auch befindet sich der letzte Eintrag in jeder Datei in der Zeile 362 zum Simulationszeitpunkt 3600 Sekunden.

Diese Dateien geben Auskunft über die detaillierte chemische Struktur der Polymere. Sie sind für die weitere Auswertung allerdings nicht relevant.

D.120.cld, D.600.cld, D.1200.cld, D.1800.cld, D.2400.cld, D.3000.cld, D.3600.cld

Diese Dateien geben Auskunft über die Kettenlängenverteilung der Polymere zu den Zeitpunkten 120, 600, 1200, 1800, 2400, 3000 und 3600 jeweils in Sekunden. Der grundsätzliche Aufbau der Dateien ist identisch.

Die Zeilen 1 bis 3 enthalten Informationen über die verwendeten Monomere.

In Zeile 4 sind die mittleren molaren Massen \overline{M}_n und \overline{M}_w sowie deren Verhältnis protokolliert.

Zeile 5 enthält Angaben über die absolute Anzahl an Polymerketten sowie Angaben über die Anzahl der SCB, die Anzahl der LCB und nochmals die Anzahl der Monomere.

Zeile 6 enthält Angaben über die Polymerisationsgrade P_n und P_w sowie die Polydispersität als deren Verhältnis.

In Zeile 7 bezeichnen „chainLength“, „nChains“, „SCB“, „LCB“, „M0“ und „M1“ die Spalten einer nachfolgenden Häufigkeitstabelle mit absoluten Werten.

log.dat, ktlog.dat

Die Datei log.dat ist die aktuelle Kopie aus dem Simulator „master“ und enthält Angaben über den Simulator selbst. Protokolliert sind unter anderem die 9 möglichen chemischen Spezies, insgesamt 11 Abbruchparameter auf Basis der Arrhenius-Gleichung sowie die insgesamt 17 möglichen Typen von Reaktionen.

Die Datei ktlog.dat ist in drei Spalten aufgeteilt. Protokolliert werden Werte für „ktss“ und „ktsm“ im Zeitverlauf nach 10 Sekunden bis zum Abbruch der Reaktion nach 3600 Sekunden. Die Zeichenkette „kt“ steht für termination rate coefficient, es handelt sich dabei um Abbruchparameter bei Kombinationsabbruch [2]. Diese werden mittels Arrhenius-Gleichungen bestimmt. Die Werte sind berechnet auf Basis der Kettenlängenverteilungen. In der Spalte „ktss“ werden die Abbruchparameter protokolliert, wenn die chemischen Spezies SPR-SPR aufeinandertreffen. Analog stehen die Werte der dritten Spalte „ktsm“ für Abbruchparameter für aufeinandertreffende Spezies SPR-MCR.

2.3 Statistische Methoden

Nachfolgend werden alle statistischen Methoden erläutert, die in der Datenauswertung mit der Statistik Software R Anwendung finden.

Modifizierter Boxplot

Ein Boxplot eignet sich dazu einen Überblick über die Lage und Streuung der vorliegenden Daten zu gewinnen. Dazu wird die Fünf-Punkte-Zusammenfassung einer Verteilung visualisiert [7, S. 59 ff.]. Diese besteht aus dem 25%- und 75%-Quantil $\tilde{x}_{0.25}$ und $\tilde{x}_{0.75}$ (der Box mit der Breite des Interquartilsabstandes $\tilde{x}_{0.75} - \tilde{x}_{0.25}$), dem Median $\tilde{x}_{0.5}$ als Linie innerhalb der Box sowie dem Minimum x_{min} und dem Maximum x_{max} .

Hierbei unterteilt ein p -Quantil die Daten in zwei Teile, sodass mindestens $p * 100\%$ der Daten darunter und $(1 - p) * 100\%$ der Daten darüber liegen [7, S. 60].

In dieser Arbeit wird der modifizierte Boxplot genutzt, bei dem Werte unterhalb $z_u = \tilde{x}_{0.25} - 1.5(\tilde{x}_{0.75} - \tilde{x}_{0.25})$ und oberhalb $z_o = \tilde{x}_{0.75} + 1.5(\tilde{x}_{0.75} - \tilde{x}_{0.25})$ als Ausreißer (einzelne Punkte) gekennzeichnet werden [7, S. 62].

In Abbildung 2.3 ist ein modifizierter Boxplot schematisch dargestellt.

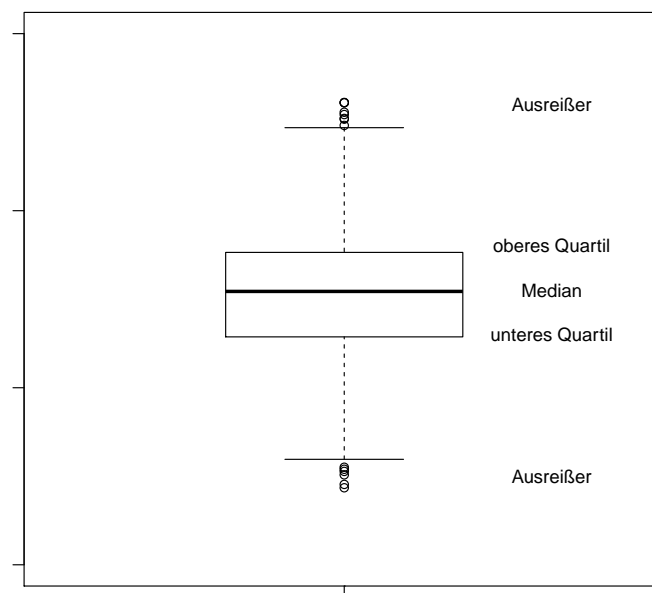


Abb. 2.3: Schematische Darstellung eines modifizierten Boxplots

Boxplots werden in R mit der Funktion `boxplot(x)` erstellt [11].

Histogramm

Ein Histogramm eignet sich dazu einen Überblick über die Gestalt einer empirischen Verteilung zu gewinnen, d.h. ob sie beispielsweise eingipfelig oder mehrgipfelig ist und ob sie symmetrisch oder schief ist.

Zur Generierung eines Histogramms muss eine Klassenbildung der Daten erfolgen. Dazu werden die Daten in benachbarte, disjunkte Intervalle $[c_0, c_1), [c_1, c_2), \dots, [c_{k-1}, c_k]$ unterteilt [7, S. 38 ff.]. Diese haben in dieser Arbeit eine feste Klassenbreite $d_j = c_j - c_{j-1}$. Durch die Klasseneinteilung ergibt sich eine unstetige Treppenfunktion.

Histogramme werden hier so konstruiert, dass die dargestellten Höhen proportional zu den relativen Häufigkeiten geteilt durch die Klassenbreite $\frac{f_j}{d_j}$ sind. Damit sind die Flächen der Rechtecke proportional zu den relativen Häufigkeiten. Dazu wird auf der vertikalen Achse die Dichte abgebildet.

In Abbildung 2.4 ist ein Histogramm schematisch dargestellt.

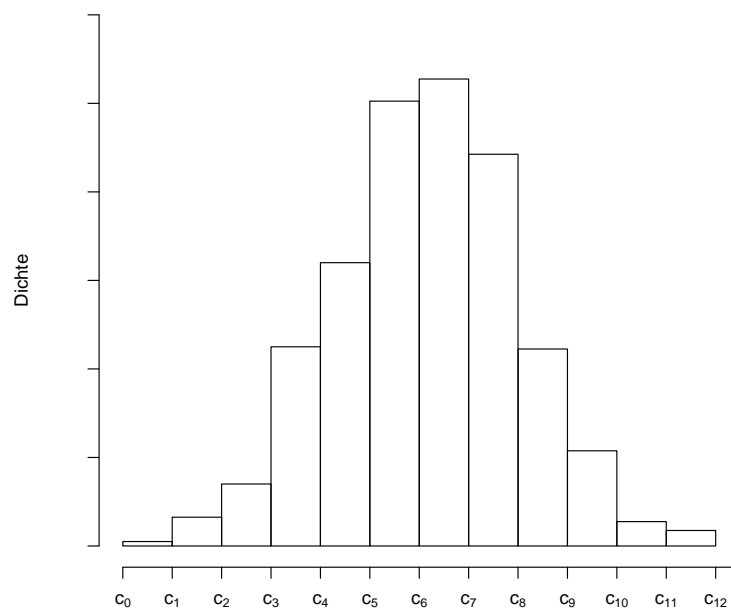


Abb. 2.4: Schematische Darstellung eines Histogramms

Histogramme werden in R mit der Funktion *hist(x)* erstellt [11].

Statistische Tests

Das Testen von Hypothesen erlaubt es, Vermutungen über wahre Parameter oder Verteilungen aus der Grundgesamtheit zu überprüfen [7, S. 369 ff.]. Dazu muss für jedes statistische Testproblem eine Nullhypothese H_0 und eine Alternative H_1 aufgestellt werden, die sich gegenseitig ausschließen. Ferner müssen die Hypothesen Aussagen über die gesamte Verteilung oder über bestimmte Parameter des interessierenden Merkmals in der Grundgesamtheit beinhalten [7, S. 384].

In dieser Arbeit werden ausschließlich zweiseitige Testprobleme angewandt, die den Erwartungswert oder die Varianz testen. Die Testprobleme lauten jeweils $H_0 : „ = “$ gegen $H_1 : „ \neq “$. Aus den statistischen Testproblemen ergibt sich eine Teststatistik, aus der sich wiederum die kritischen Bereiche ermitteln lassen, wann die Nullhypothese H_0 zu verwerfen ist.

Als Beispiel dient ein Münzwurf. Dieser ist binominalverteilt und hat die Erfolgswahrscheinlichkeit p . Bei einer fairen Münze würde man davon ausgehen, dass mit gleicher Wahrscheinlichkeit Kopf oder Zahl eintritt, dass heißt dass der Parameter der Erfolgswahrscheinlichkeit (beispielsweise mit Kopf als Erfolg) gleich 0.5 ist. Entsprechend kann die Nullhypothese $H_0 : p = 0.5$ formuliert werden. Die Hypothese, dass eine unverfälschte Münze vorliegt, wird nur dann verworfen, wenn einer der zum Signifikanzniveau ermittelten kritischen Wahrscheinlichkeitswerte unterschritten bzw. überschritten wird.

Weil man mit der Durchführung eines statistischen Tests von einer konkreten Stichprobe auf die Grundgesamtheit schließt sind Fehlentscheidungen nie ganz ausschließbar [7, S. 385]. Es gibt zwei Arten von Fehlentscheidungen.

Als Fehler 1. Art gilt, wenn H_0 verworfen wird, obwohl H_0 wahr ist.

Als Fehler 2. Art gilt, wenn H_0 nicht verworfen wird, obwohl H_1 wahr ist.

Im Allgemeinen ist eine gleichzeitige Minimierung beider Fehlerarten nicht möglich. Daher versucht man vorrangig den Fehler 1. Art durch eine obere Schranke zu kontrollieren. Ein statistischer Test zum Signifikanzniveau α begrenzt die Wahrscheinlichkeit des ersten Fehlers auf $P(H_1 \text{ annehmen} | H_0 \text{ wahr}) \leq \alpha$.

Das Signifikanzniveau α wird typischerweise auf 0.1, 0.05 oder 0.01 festgelegt.

Weiterhin sollte beachtet werden, dass an jedes statistische Testproblem Annahmen geknüpft sind, die erfüllt sein müssen. Die Annahmen werden im Abschnitt des jeweiligen Tests gesondert beschrieben.

p-Wert

Der p-Wert stellt eine Überschreitungswahrscheinlichkeit dar und wird bei der Anwendung von statistischen Tests in R automatisch ausgegeben [7, S. 387 ff.]. Er kann Werte größer gleich 0 und kleiner gleich 1 annehmen.

Definiert ist der p-Wert als die Wahrscheinlichkeit, „unter H_0 den beobachteten Prüfgrößenwert oder einen in Richtung der Alternative extremeren Wert zu erhalten.“ [7, S. 388]

Die Nullhypothese H_0 ist immer dann zu verwerfen, wenn der p-Wert kleiner als das vorab festzulegende Signifikanzniveau α ist. Andernfalls wird die Nullhypothese H_0 nicht verworfen.

Konfidenzintervall für den Erwartungswert

Konfidenzintervalle stellen eine Unsicherheitsangabe dar und sind das Ergebnis einer Intervallschätzung [7, S. 356 ff.]. Sie sind allgemein mit $[G_u, G_o]$ durch eine untere und eine obere Intervallgrenze definiert. Dabei sind die Intervallgrenzen Zufallsgrößen und keine deterministischen Werte. Das Intervall ist so konstruiert, dass es mit einer Wahrscheinlichkeit von $1 - \alpha$ den wahren Parameter überdeckt.

Konfidenzintervalle gibt es für eine Vielzahl von Parametern wie den Erwartungswert μ oder die Varianz σ^2 . Von Interesse ist hier lediglich das Konfidenzintervall für den Erwartungswert μ . Der Lageparameter soll auf einen konkreten Wert μ_0 getestet werden. Für den Ein-Stichproben-t-Test lautet die Nullhypothese $H_0 : \mu = \mu_0$ und die Alternative $H_1 : \mu \neq \mu_0$.

Eine Möglichkeit das Konfidenzintervall herzuleiten besteht darin, die kritischen Werte des Ein-Stichproben-t-Tests zu verwenden. Dazu wird die Teststatistik des Ein-Stichproben-t-Tests in Anlehnung an [7, S. 401] verwendet.

$$T = \frac{\bar{X} - \mu_0}{S} \sqrt{n} \quad (2.8)$$

Der Test setzt voraus, dass die Daten normalverteilt sind [12, S. 95 f.]. Die Realisationen der unabhängigen Zufallsvariablen X_1, \dots, X_n lauten x_1, \dots, x_n . \bar{X} ist das Stichprobenmittel. Die Stichprobengröße wird durch die Variable n ausgedrückt.

Da σ^2 beim t-Test unbekannt ist, wird mit S ein Schätzer für σ verwendet.

$$S = \sqrt{\frac{1}{n-1} \sum_i (X_i - \bar{X})^2} \quad (2.9)$$

Daraus ergibt sich das exakte $(1 - \alpha)$ -Konfidenzintervall für μ bei normalverteiltem Merkmal in Anlehnung an [7, S. 360] wie folgt.

$$\left[\bar{X} - t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}} \right] \quad (2.10)$$

Hierbei ist $t_{1-\alpha/2}(n-1)$ das $(1 - \alpha/2)$ -Quantil der t -Verteilung mit $n - 1$ Freiheitsgraden.

Das hier beschriebene Konfidenzintervall kann in R automatisch erstellt werden indem man einen Ein-Stichproben- t -Test mit $t.test(x)$ anwendet [11].

Zwei-Stichproben- t -Test

Ein Zwei-Stichproben- t -Test prüft, ob sich die Erwartungswerte zweier Stichproben bedeutend unterscheiden.

Es existieren zwei Stichproben X_1, \dots, X_n unabhängig und identisch verteilt wie X und Y_1, \dots, Y_m unabhängig und identisch verteilt wie Y . Die Stichprobengrößen n von X und m von Y sind nicht notwendigerweise gleich groß. Der Erwartungswert von X ist μ_X , der Erwartungswert von Y ist μ_Y . \bar{X} ist das Stichprobenmittel von X und \bar{Y} das Stichprobenmittel von Y .

Es wird die Nullhypothese $H_0 : \mu_X - \mu_Y = \delta_0$ gegen die Hypothese $H_1 : \mu_X - \mu_Y \neq \delta_0$ getestet [7, S. 419]. Mit $\delta_0 = 0$ entspricht dies der Fragestellung, dass die Erwartungswerte gleich sind.

Für den Zwei-Stichproben- t -Tests liegt in Anlehnung an [7, S. 420] folgende Teststatistik vor.

$$T = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \quad (2.11)$$

Da auch hier die Varianzen unbekannt sind, werden mit S_X^2 und S_Y^2 Schätzer verwendet.

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S_Y^2 = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2 \quad (2.12)$$

In den Testergebnissen von R sind die Unterschiede der Erwartungswerte als ein

Konfidenzintervall angegeben. Bei Verwerfen der Nullhypothese ist darin keine Null enthalten [12, S. 102].

Zwei-Stichproben-t-Tests werden in R mit der Funktion `t.test(x,y)` durchgeführt [11].

Varianz-Test

Ein Varianz-Test prüft, ob sich die Varianzen zweier Stichproben bedeutend unterscheiden [12, S. 103]. Es existieren zwei Zufallsvariablen X und Y . Der Test setzt voraus, dass alle Beobachtungen aus beiden Stichproben unabhängig voneinander sind. Zudem müssen die Realisationen x_1, \dots, x_n aus einer Normalverteilung mit den Parametern Erwartungswert μ_1 und Standardabweichung σ_1 stammen und die Realisationen y_1, \dots, y_n analog aus einer Normalverteilung mit den Parametern μ_2 und σ_2 [13, S. 158].

Es wird die Nullhypothese $H_0 : \sigma_1^2/\sigma_2^2 = \gamma$ gegen die Hypothese $H_1 : \sigma_1^2/\sigma_2^2 \neq \gamma$ mit $\gamma = 1$ getestet [13, S. 158].

Die Teststatistik lautet in Anlehnung an [13, S. 158] wie folgt.

$$F = \frac{S_X^2}{S_Y^2} \quad (2.13)$$

Da auch hier die Varianzen unbekannt sind, werden mit S_X^2 und S_Y^2 die schon bekannten Schätzer verwendet.

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S_Y^2 = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2 \quad (2.14)$$

In den Ausgaben der Tests wird außerdem das Verhältnis der Varianzen angegeben. Varianz-Tests werden in R mit der Funktion `var.test(x,y)` erstellt [11].

Kern-Dichteschätzer

Ein Kern-Dichteschätzer ist ein Schätzer für Dichtekurven [7, S. 93]. Mit einem stetigem Kern kann dieser den Nachteil von Histogrammen ausgleichen, dass durch die Gruppierung der Daten in Klassen eine unstetige Treppenfunktion abgebildet wird [7, S. 80 ff.]. Der Kern-Dichteschätzer wird wie folgt berechnet.

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - x_i}{h}\right) \quad (2.15)$$

In dieser Arbeit wird der Gauß-Kern angewandt.

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right), u \in \mathbb{R} \quad (2.16)$$

Die Berechnung einer approximierten Dichtekurve kann in R mit der Funktion *density(x)* erreicht [13, S. 69]. Ein wichtiger Parameter der Funktion ist die Bandbreite h , die die Glattheit der einzuziehenden Kurve kontrolliert. Sie wird in R aus der Multiplikation der Argumente *bw* und *adjust* berechnet, wobei letzteres ein einfacher Faktor ist.

Das Argument *bw* wird in Anlehnung an [14, S. 127] in R wie folgt berechnet.

$$h = 0.9 \min\{S_x, (\tilde{x}_{0.75} - \tilde{x}_{0.25})/1.34\} n^{-1/5} \quad (2.17)$$

Wird die Bandbreite zu schmal gewählt, können unwichtige Funktionsteile hervortreten. Wird sie zu breit gewählt, werden möglicherweise wichtige Funktionsteile ausgeglättet. Dies wird anhand Abbildung 2.5 deutlich.

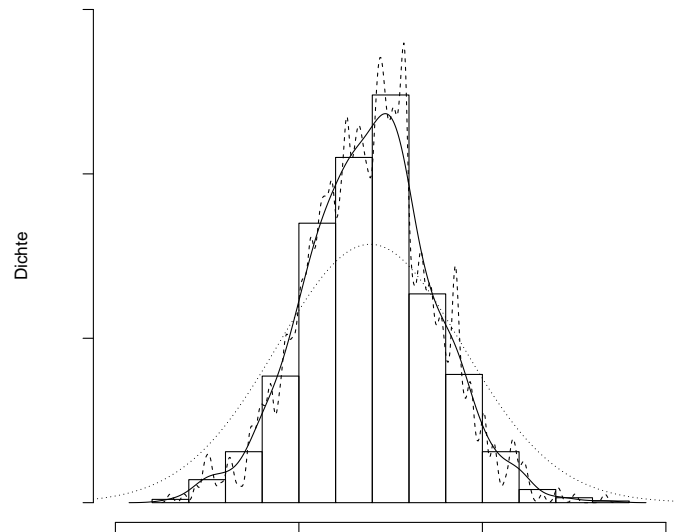


Abb. 2.5: Darstellung von Dichtekurven in einem Histogramm. Die Werte für *adjust* sind 0.2 (gestrichelte Linie), 1 (durchgehende Linie) und 5 (gepunktete Linie)

Wie man sieht ist die Glättung mit einem Wert für *adjust* von 0.2 nicht stark genug. Es treten viele Funktionsteile zu Tage, die das Histogramm selbst gar nicht zeigt. Ebenso ist die Glättung mit einem Wert für *adjust* von 5 viel zu stark, da die Funktion nicht detailliert genug wiedergegeben wird. Ein sehr guter Kompromiss aus Glättung und exakter Wiedergabe aller Spitzen der Daten zeigt die durchgehende Linie mit einem Wert für *adjust* von 1.

Das Einzeichnen der Dichtekurve in ein Histogramm erfolgt mit der Funktion *lines(x)* [14, S. 73]. In dieser Arbeit wird die Standardeinstellung der Bandbreite verwendet.

3 Explorative Datenanalyse

In den vorhergehenden Abschnitten 2.2.2 und 2.3 konnte bereits ein Überblick über die Primärdaten und die zur Anwendung kommenden statistischen Methoden gewonnen werden. Nun werden die vorhandenen Simulationsergebnisse explorativ ausgewertet.

Aus praktischen Gründen erfolgt die Datenauswertung ab einem Zeitpunkt von 600 Sekunden, da erst hier unter den gewählten Reaktionsbedingungen eine ausreichend große Menge Polymer gebildet wurde.

3.1 Auswertung einer einzelnen Simulation

Die Auswertung einer einzelnen Simulation über die Zeit erfolgt exemplarisch anhand der Daten im Ordner `sim1` für $1 \cdot 10^{11}$ Moleküle.

Für das Einlesen der Werte wird der Befehl `read.table()` genutzt [11]. Um nur die absoluten Häufigkeiten einzulesen wird dabei die Option `skip=6` verwendet. Die Ausgabe der Grafiken erfolgt mit dem Befehl `plot(x)` [11].

In den Abbildungen 3.1 und 3.2 sind die absoluten Häufigkeiten „nChains“ der Kettenlängen „chainLength“ für den zweiten und letzten protokollierten Simulationszeitpunkt dargestellt. Die Skalen der x-Achse sind identisch, während sich die Werte der y-Achse im Zeitverlauf um mehr als Faktor 6 erhöhen. Ein Blick in die Kettenlängenverteilungen `D.600.0.cld` und `D.3600.0.cld` verrät, dass im Polymerisat nach 600 Sekunden rund 2.5 Milliarden und nach einer Stunde etwa acht Milliarden Monomere enthalten sind.

Die Verteilung der molaren Masse in Abbildung 3.3 erhält man schließlich durch Multiplikation der Werte der Abzisse mit der molaren Masse des Monomers. Es handelt sich bei der Verteilung um eine Funktion diskreter Art, da die beobachtbaren Werte endlich abzählbar sind [7, S. 15]. Es ergibt sich für die radikalische Polymerisation eine unimodale, rechtsschiefe Verteilung.

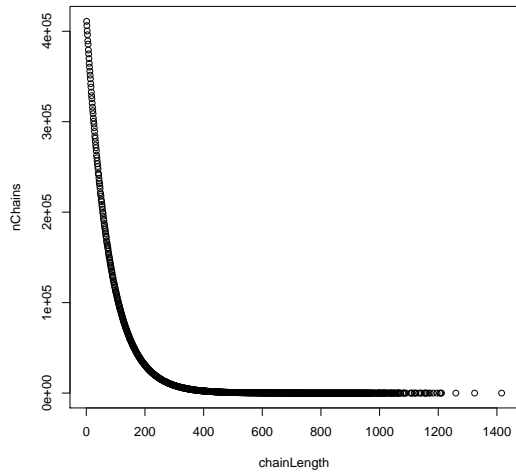


Abb. 3.1: Kettenlängenverteilung für die Simulation sim1 nach 600 Sekunden

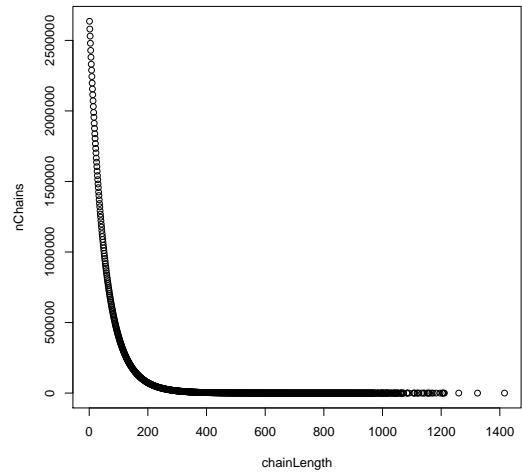


Abb. 3.2: Kettenlängenverteilung für die Simulation sim1 nach 3600 Sekunden

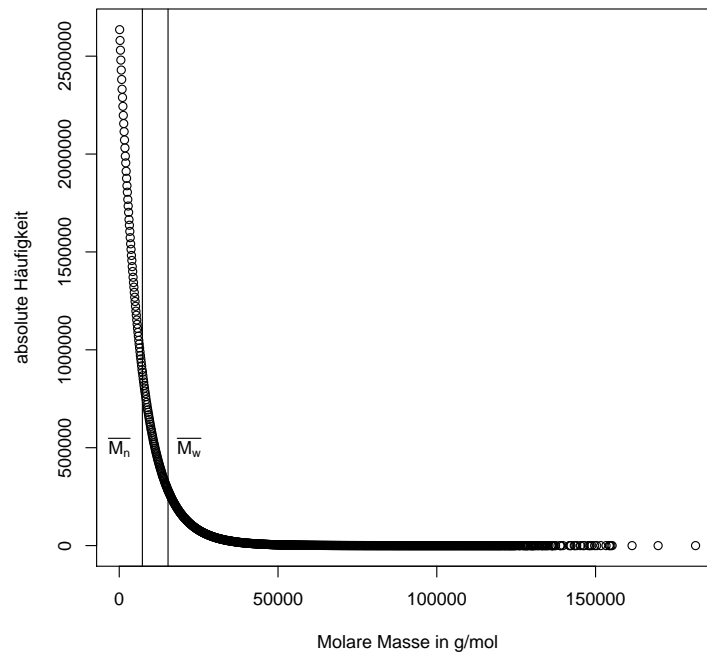


Abb. 3.3: Verteilung der molaren Masse für die Simulation sim1 nach 3600 Sekunden. In der Grafik eingezeichnet sind das Zahlenmittel $\overline{M}_{n,3600}$ und das Gewichtsmittel $\overline{M}_{w,3600}$

Auch wenn durch diese Anwendung ein Informationsverlust einhergeht werden die Häufigkeitsverteilungen durch die beiden Lagemaße Zahlenmittel \overline{M}_n und Gewichtsmittel \overline{M}_w substituiert. Dies erlaubt eine vereinfachte Datenauswertung.

Es ergeben sich gerundet die Zahlenmittel $\overline{M}_{n,600} = 9932.02$ und $\overline{M}_{n,3600} = 7287.53$ sowie die beiden Gewichtsmittel $\overline{M}_{w,600} = 19792.93$ und $\overline{M}_{w,3600} = 15370.76$. Dies entspricht einer Abnahme der Zahlenwerte über die Zeit von jeweils etwa 25 Prozent. Der Verlauf der Werte ist in Abbildung 3.4 dargestellt.

Der Umsatz des Monomers X_t wird entsprechend der Formel im Abschnitt 2.1.3 berechnet. Dieser steigt stetig von einem Wert von 0.184 zum Zeitpunkt 600 Sekunden auf 0.596 zum Zeitpunkt 3600 Sekunden an. Der Verlauf des Umsatzes des Monomers ist in Abbildung 3.5 dargestellt.

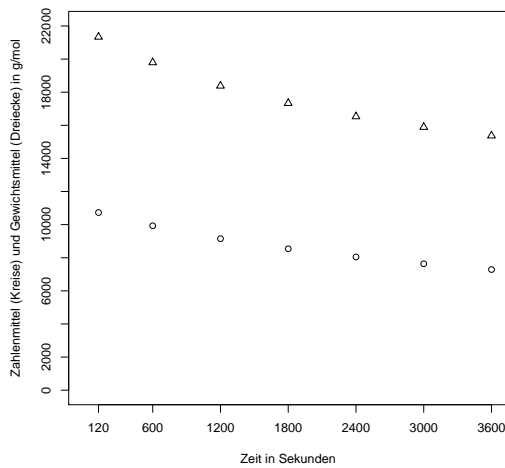


Abb. 3.4: Entwicklung von \overline{M}_n und \overline{M}_w für die Simulation sim1

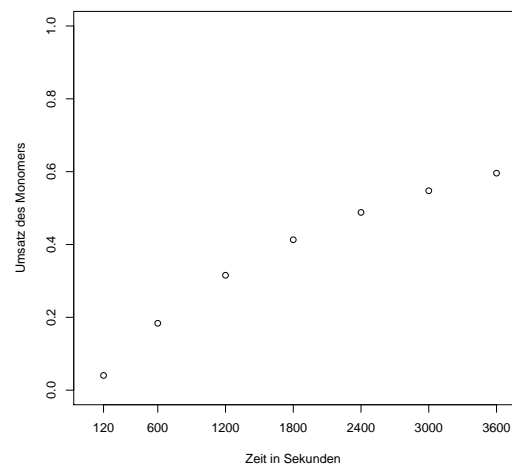


Abb. 3.5: Verlauf des Umsatzes des Monomers X_t für die Simulation sim1

Die beschriebenen Ergebnisse waren schon vor der Bearbeitung bekannt und wurden von Drache et al. publiziert [2].

3.2 Auswertung innerhalb der Settings

Zunächst werden die sieben Simulationszeitpunkte als Vektor in der Variable *time* hinterlegt. Anschließend werden die sieben Dateinamen innerhalb einer Simulation

mittels dem Befehl *paste()* [11] und einer *for-Schleife* generiert. Schließlich werden mittels zwei verschachtelten *for-Schleifen*, der Funktion *read.table()* und einer weiteren Anwendung des Befehls *paste()* die Daten den Variablen *e11.MnMw*, *e10.MnMw*, *e11.c* und *e10.c* zugewiesen und als Matrix gespeichert, um auf die Werte spaltenweise zugreifen zu können. Die Daten werden als .csv-Datei exportiert und anschließend wieder importiert, damit die Algorithmen nur einmal durchlaufen werden müssen.

Als finale Datensätze ergeben sich für *e11.MnMw* und *e11.c* eine 1000x14-Matrix und für *e10.MnMw* und *e10.c* eine 2673x14-Matrix.

Die Auswertung erfolgt über alle Simulationen hinweg und getrennt nach der Molekülanzahl. Die Zusammenhänge sollen durch die Generierung von Boxplots besser aufgedeckt werden.

Auswertung für $1 * 10^{11}$ Moleküle (graue Boxplots)

Die grauen Boxplots in den Abbildungen 3.6 und 3.7 stellen die Streuung der beiden Lagemaße der molaren Masse \overline{M}_n und \overline{M}_w für eine Molekülanzahl von $1 * 10^{11}$ zu den drei Simulationszeitpunkten 600 Sekunden, 1800 Sekunden und 3600 Sekunden dar.

Es sind eindeutig abnehmende Zahlenwerte für die beiden Lagemaße \overline{M}_n und \overline{M}_w im Zeitverlauf feststellbar. Auch sinken deren Interquartilabstände und Varianzen.

Der Umsatz des Monomers X_t steigt stetig und weist wie in Abbildung 3.8 ersichtlich nur eine geringe Streuung auf. Dessen Median beträgt 0.184 zum Zeitpunkt 600 Sekunden und 0.596 zum Zeitpunkt 3600 Sekunden.

Die in Abschnitt 3.1 beschreibenden Trends werden gut wiedergegeben. Die Beobachtungen in den einzelnen Boxplots haben kaum einen Einfluss auf das Gesamtergebnis.

Auswertung für $5 * 10^{10}$ Moleküle (weiße Boxplots)

Die abnehmenden Zahlenwerte im Zeitverlauf sind auch bei einer Molekülanzahl von $5 * 10^{10}$ für beide Lagemaße \overline{M}_n und \overline{M}_w eindeutig feststellbar. Ebenso sinken deren Interquartilabstände und Varianzen.

Der Umsatz des Monomers X_t steigt auch hier stetig und beläuft sich auf die Mediane 0.184 zum Zeitpunkt 600 Sekunden und 0.597 zum Zeitpunkt 3600 Sekunden.

Die in Abschnitt 3.1 beschreibenden Trends werden gut wiedergegeben. Die Beobachtungen in den einzelnen Boxplots haben kaum einen Einfluss auf das Gesamtergebnis.

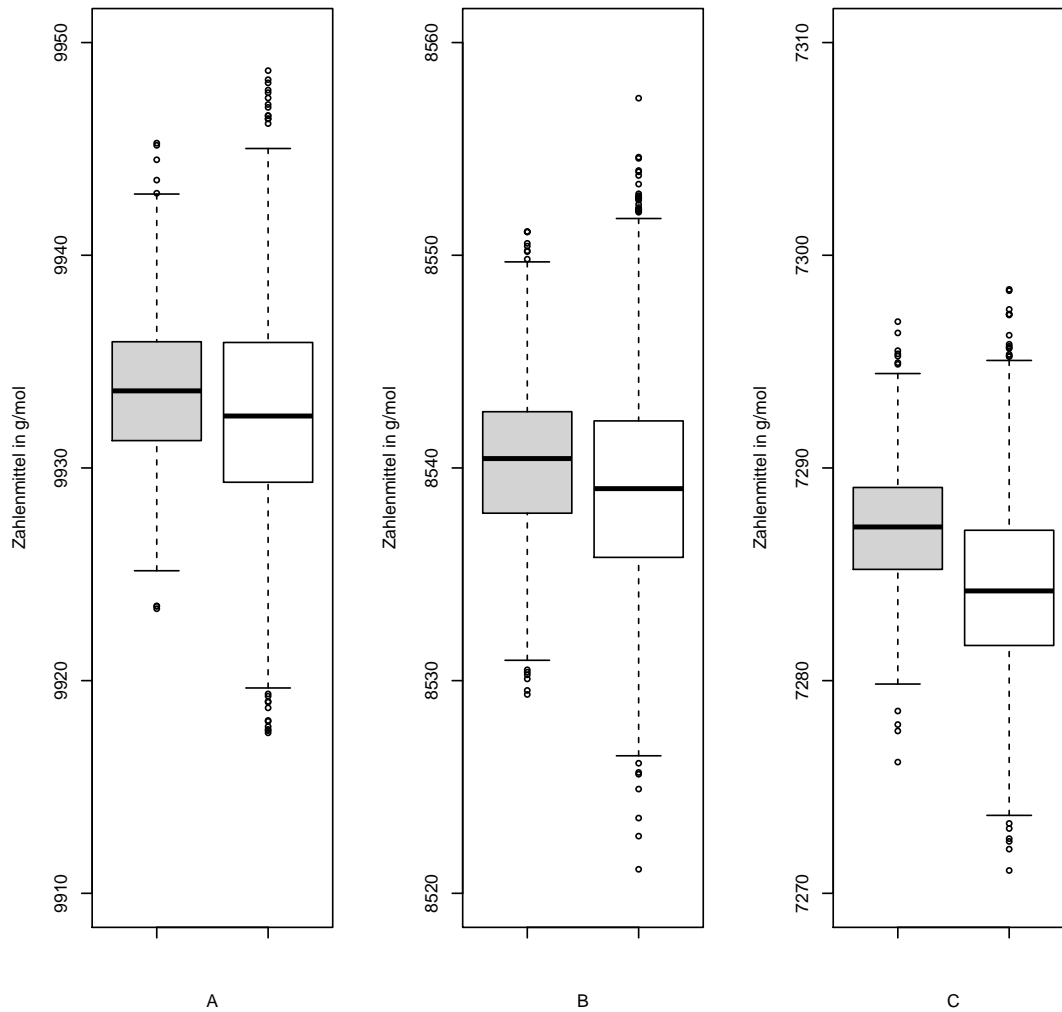


Abb. 3.6: Boxplots des Zahlenmittels \overline{M}_n für jeweils $1 \cdot 10^{11}$ Moleküle (grau) und $5 \cdot 10^{10}$ Moleküle (weiß) nach 600 Sekunden (A), 1800 Sekunden (B) und 3600 Sekunden (C)

3.3 Auswertung zwischen den Settings

Die Auswertung erfolgt wieder über alles Simulationen hinweg.

Zwischen den Settings mit $1 \cdot 10^{11}$ Molekülen und $5 \cdot 10^{10}$ Molekülen ist auffällig, dass die Mediane der mittleren molaren Massen \overline{M}_n und \overline{M}_w für $1 \cdot 10^{11}$ Moleküle in allen Fällen einen höheren Wert annehmen. Die Varianzen und Interquartilabstände fallen bei größerer Molekülanzahl zudem in allen Fällen deutlich geringer aus. Außerdem

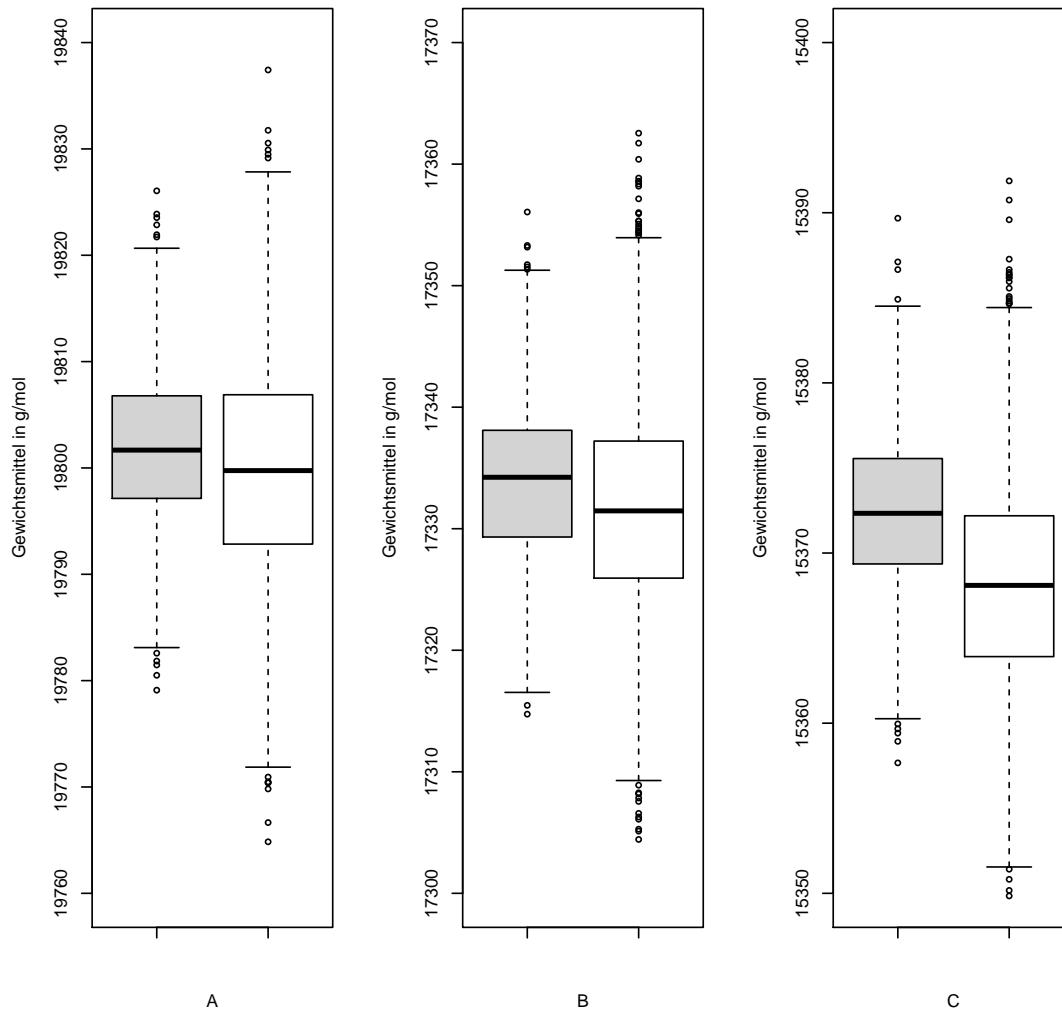


Abb. 3.7: Boxplots des Gewichtsmittels \overline{M}_w für jeweils $1 \cdot 10^{11}$ Moleküle (grau) und $5 \cdot 10^{10}$ Moleküle (weiß) nach 600 Sekunden (A), 1800 Sekunden (B) und 3600 Sekunden (C)

enthalten die Boxplots für eine Molekülanzahl von $5 \cdot 10^{10}$ mehr Ausreißer.

Der Umsatz des Monomers X_t weist bei einer Molekülanzahl von $1 \cdot 10^{11}$ einen kleineren Median und eine geringere Streuung auf als bei einer Molekülanzahl von $5 \cdot 10^{10}$. Auch hier können für die kleinere Molekülanzahl von $5 \cdot 10^{10}$ mehr Ausreißer verzeichnet werden.

Es kann festgestellt werden, dass alle drei Zielgrößen \overline{M}_n , \overline{M}_w und Umsatz X_t bezüglich der Streuung und den Ausreißern gleichartig betroffen sind.

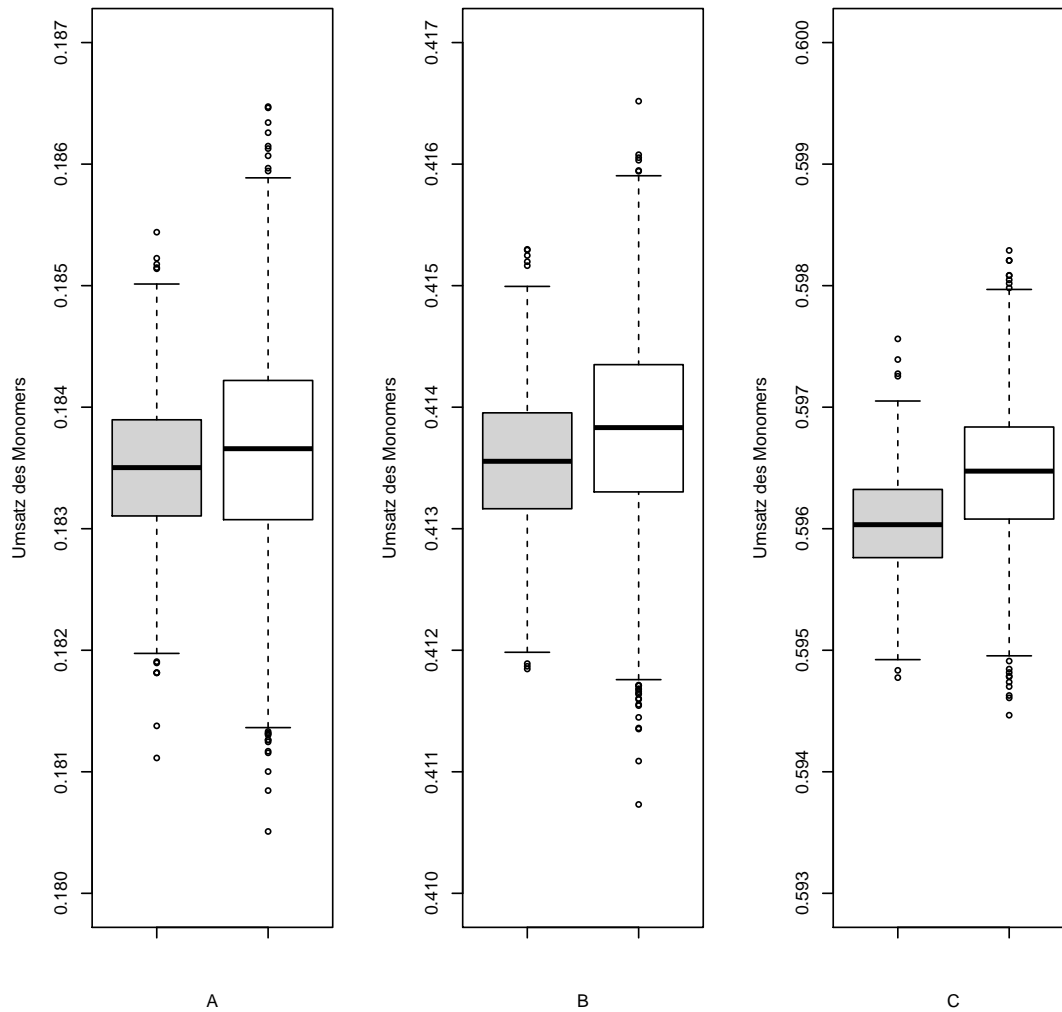


Abb. 3.8: Boxplots des Umsatzes des Monomers X_t für jeweils $1 \cdot 10^{11}$ Moleküle (grau) und $5 \cdot 10^{10}$ Moleküle (weiß) nach 600 Sekunden (A), 1800 Sekunden (B) und 3600 Sekunden (C)

Die Unterschiede der Mediane sind mit $M_{BuA} = 128.17 \frac{\text{g}}{\text{mol}}$ in der Größenordnung von einem Hundertstel eines Monomers (vgl. Formel 2.4). Alle Schwankungen sind viel kleiner als eine Monomereinheit.

Weiterhin sind die Simulationsergebnisse viel präziser als Experimente (siehe [2] Abbildung 2), wo Abweichungen von $1000 \frac{\text{g}}{\text{mol}}$ noch akzeptiert werden.

Diese Erkenntnisse können mit Histogrammen alternativ visualisiert werden.

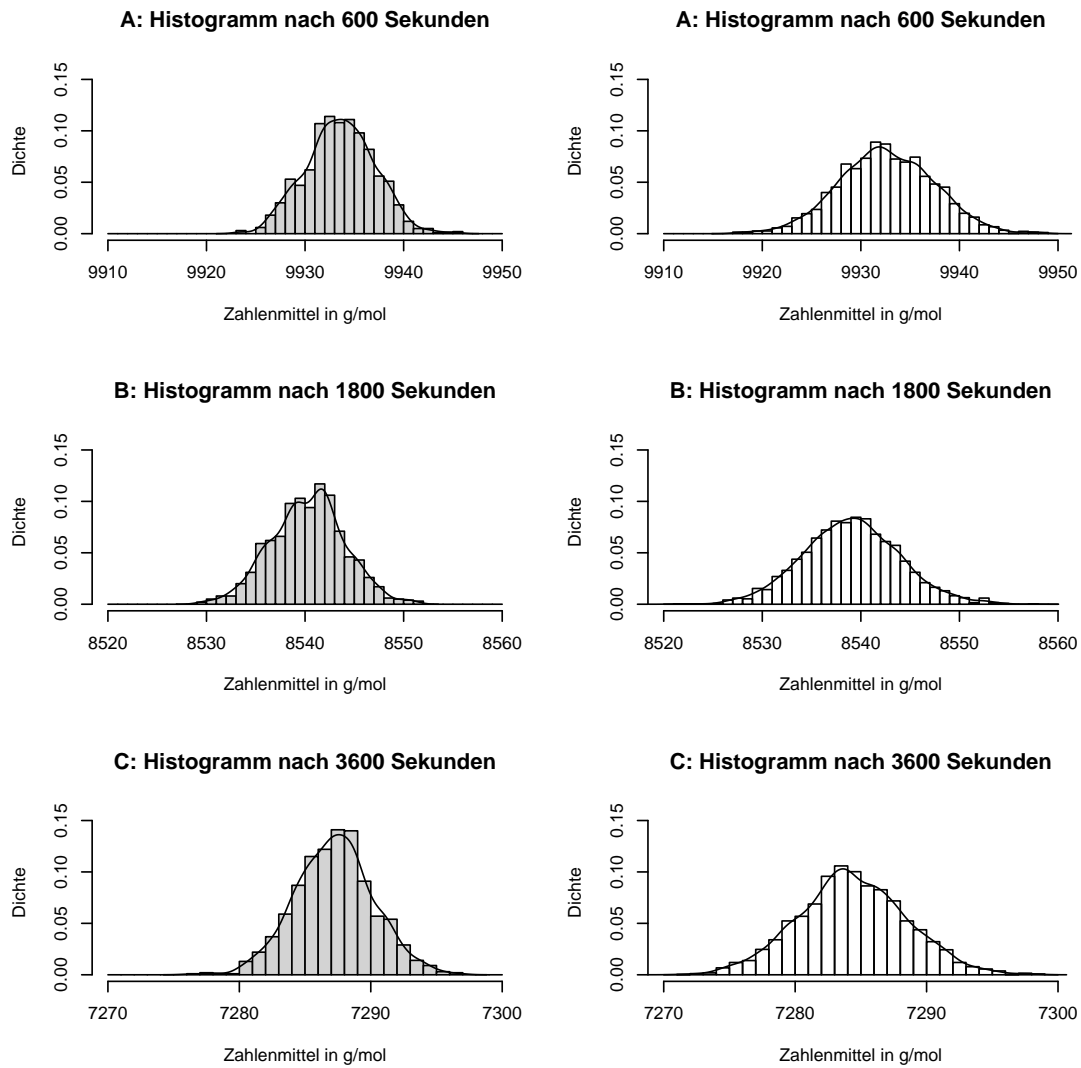


Abb. 3.9: Histogramme des Zahlenmittels \overline{M}_n für jeweils $1 \cdot 10^{11}$ Moleküle (grau) und $5 \cdot 10^{10}$ Moleküle (weiß) nach 600 Sekunden (A), 1800 Sekunden (B) und 3600 Sekunden (C)

Die Histogramme in Abbildung 3.9 und 3.10 belegen für das Zahlenmittel \overline{M}_n und das Gewichtsmittel \overline{M}_w , dass bei einer Molekülanzahl von $1 \cdot 10^{11}$ engere Verteilungen entstehen als bei einer Molekülanzahl von $5 \cdot 10^{10}$. Zudem kann erneut festgestellt werden, dass die Mediane der mittleren molaren Massen \overline{M}_n und \overline{M}_w bei einer Molekülanzahl von $1 \cdot 10^{11}$ etwas größer sind.

Die bisher gemachten Beobachtungen für den Umsatz des Monomers X_t können ebenfalls durch die Erstellung der Histogramme in Abbildung 3.11 bestätigt werden.

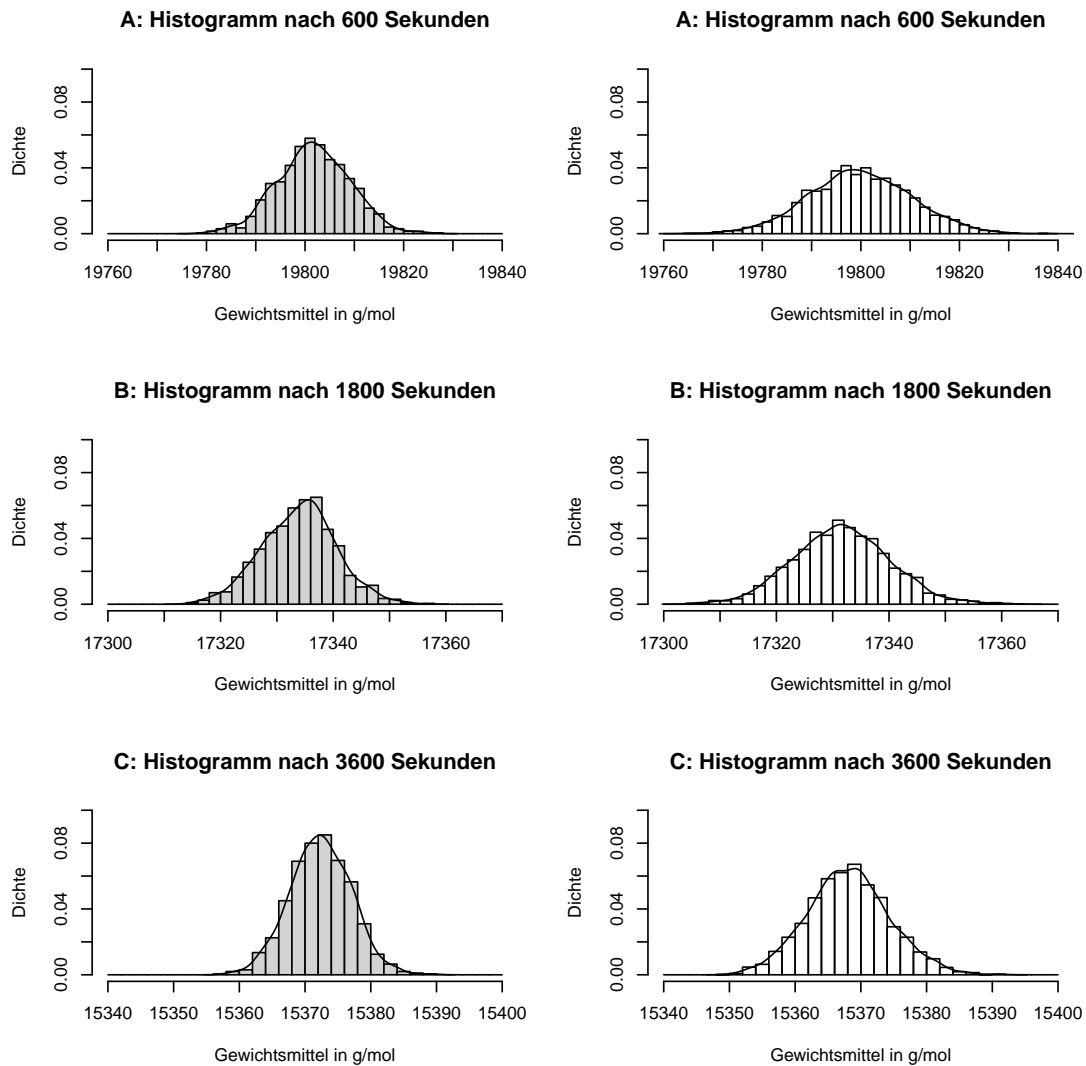


Abb. 3.10: Histogramme des Gewichtsmittels \overline{M}_w für jeweils $1 \cdot 10^{11}$ Moleküle (grau) und $5 \cdot 10^{10}$ Moleküle (weiß) nach 600 Sekunden (A), 1800 Sekunden (B) und 3600 Sekunden (C)

Bei größerer Molekülanzahl fällt die Streuung der Werte geringer aus und der Median ist bei größerer Molekülanzahl etwas kleiner.

Sowohl für die Lagemaße der molaren Masse als auch den Umsatz des Monomers kann zudem festgestellt werden, dass die Verteilungen im Zeitverlauf symmetrischer werden. Dies ist darauf zurückzuführen, dass sich im Zeitverlauf mehr Polymere bilden.

Für das Zahlenmittel \overline{M}_n , das Gewichtsmittel \overline{M}_w und den Umsatz des Monomers X_t wurden zudem Kern-Dichteschätzer in die Histogramme eingezeichnet. Diese bestätigen

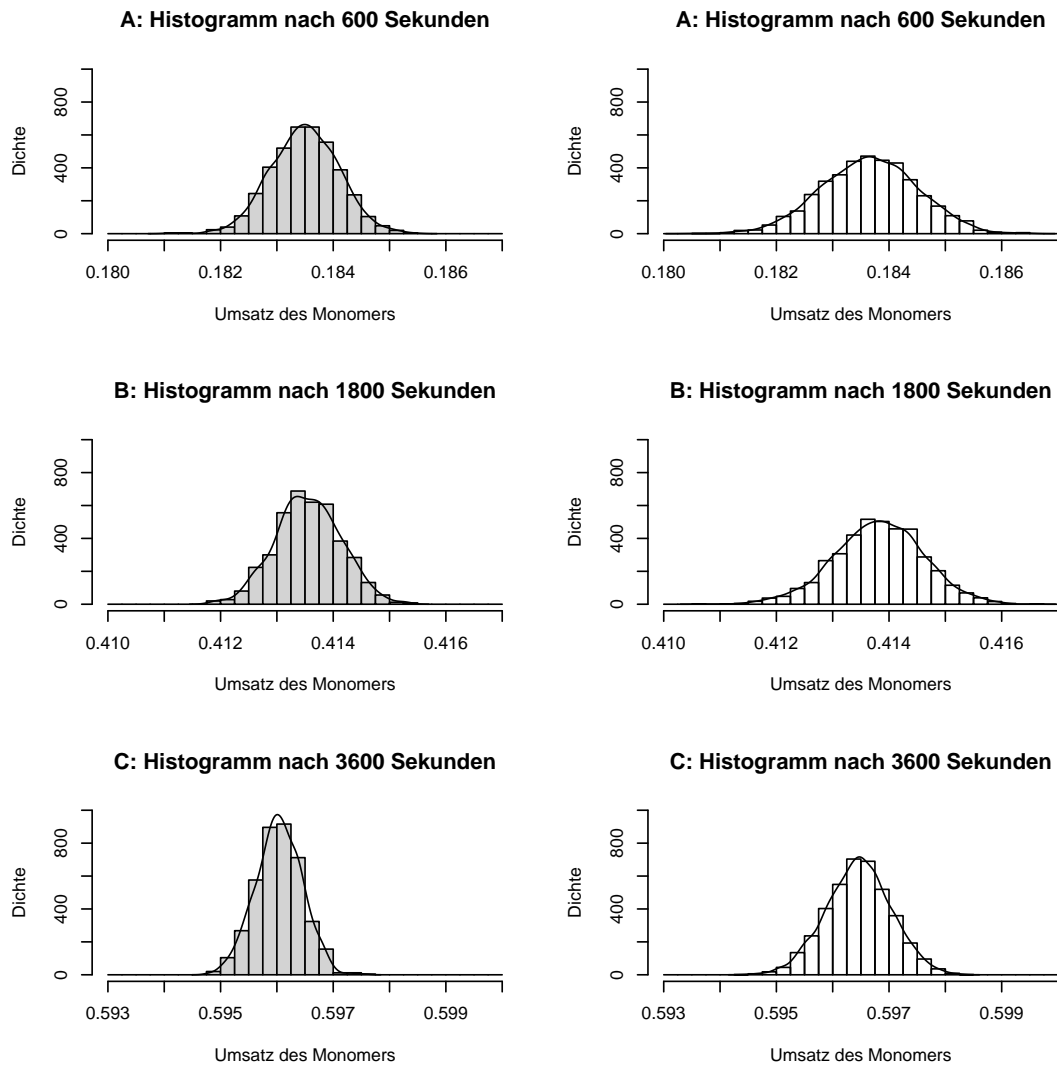


Abb. 3.11: Histogramme des Umsatzes des Monomers X_t für jeweils $1 \cdot 10^{11}$ Moleküle (grau) und $5 \cdot 10^{10}$ Moleküle (weiß) nach 600 Sekunden (A), 1800 Sekunden (B) und 3600 Sekunden (C)

den Eindruck, dass die Verteilungen im Zeitverlauf symmetrischer werden.

4 Statistische Verfahren

Aufbauend auf der explorativen Datenanalyse sollen die Erkenntnisse durch statistische Tests gestützt werden. Dazu wird das Signifikanzniveau α auf 0.05 festgelegt.

Die folgende Auswertung konzentriert sich auf die Unterschiede der beiden Lagemaße der molaren Masse \overline{M}_n und \overline{M}_w und des Umsatzes des Monomers X_i zwischen den beiden Settings und erfolgt erneut über alle Simulationen hinweg.

Die beiden Stichproben für $1 * 10^{11}$ Moleküle mit 2673 Simulationen und $5 * 10^{10}$ Moleküle mit 1000 Simulationen unterscheiden sich in der Größe so extrem, dass es sinnvoll ist, bei der größeren Stichprobe eine Teilstichprobe zu ziehen. Daher werden aus den Daten für $5 * 10^{10}$ Moleküle 1000 Simulationen ohne Zurücklegen randomisiert gezogen. Zur Reproduzierbarkeit der Stichprobe wird in diesen Fällen vorab jeweils ein Seed in R gesetzt [7, S. 307].

4.1 Konfidenzintervalle für den empirischen Mittelwert

Standardmäßig werden Ein-Stichproben-t-Tests angewandt um die Nullhypothese zu prüfen, ob der empirische Mittelwert einer Stichprobe einen bestimmten Wert annimmt. Der p-Wert ist in allen zwölf Fällen geringer als das festgelegte Signifikanzniveau. Damit wird jeweils die Nullhypothese verworfen, dass die empirischen Mittelwerte null sind.

Tatsächlich von Interesse sind hier aber die Konfidenzintervalle für den empirischen Mittelwert, die automatisch bei Anwendung der Tests in R ausgegeben werden.

Anhand der Testergebnisse in Tabelle 4.1 kann festgestellt werden, dass bei einer Molekülanzahl von $1 * 10^{11}$ die Lagemaße der molaren Masse \overline{M}_n und \overline{M}_w im Mittel etwas größer und die Konfidenzintervalle etwas schmaler sind. Diese Beobachtungen decken sich mit den Erkenntnissen aus Abschnitt 3.3.

Die Konfidenzintervalle haben eine hohe Präzision und sind besser als experimentelle Ergebnisse. Sie geben Hinweise darauf, dass sich die empirischen Mittelwerte unterscheiden.

Tab. 4.1: Konfidenzintervalle für die beiden Lagemaße

Testgröße	emp. Mittelwert	95%-Konfidenzintervall
$1 * 10^{11} \overline{M_{n,600}}$	9933.57	[9933.35, 9933.79]
$5 * 10^{10} \overline{M_{n,600}}$	9932.49	[9932.19, 9932.78]
$1 * 10^{11} \overline{M_{n,1800}}$	8540.30	[8540.07, 8540.53]
$5 * 10^{10} \overline{M_{n,1800}}$	8538.71	[8538.40, 8539.02]
$1 * 10^{11} \overline{M_{n,3600}}$	7287.23	[7287.04, 7287.41]
$5 * 10^{10} \overline{M_{n,3600}}$	7284.40	[7284.14, 7284.65]
$1 * 10^{11} \overline{M_{w,600}}$	19801.72	[19801.27, 19802.18]
$5 * 10^{10} \overline{M_{w,600}}$	19799.86	[19799.21, 19800.51]
$1 * 10^{11} \overline{M_{w,1800}}$	17333.93	[17333.52, 17334.34]
$5 * 10^{10} \overline{M_{w,1800}}$	17331.67	[17331.13, 17332.22]
$1 * 10^{11} \overline{M_{w,3600}}$	15372.36	[15372.07, 15372.65]
$5 * 10^{10} \overline{M_{w,3600}}$	15368.05	[15367.66, 15368.43]

Aus den Ausgaben der Tests in Tabelle 4.2 kann für den Umsatz des Monomers X_t festgestellt werden, dass bei der größeren Anzahl an Molekülen von $1 * 10^{11}$ die empirischen Mittelwerte etwas kleiner und die Konfidenzintervalle etwas schmaler sind. Die Erkenntnisse decken sich ebenfalls mit den Beobachtungen aus Abschnitt 3.3.

Tab. 4.2: Konfidenzintervalle für den Umsatz des Monomers

Testgröße	emp. Mittelwert	95%-Konfidenzintervall
$1 * 10^{11} X_{t,600}$	0.18350	[0.18346, 0.18353]
$5 * 10^{10} X_{t,600}$	0.18365	[0.18360, 0.18371]
$1 * 10^{11} X_{t,1800}$	0.41356	[0.41352, 0.41360]
$5 * 10^{10} X_{t,1800}$	0.41382	[0.41377, 0.41387]
$1 * 10^{11} X_{t,3600}$	0.59604	[0.59601, 0.59606]
$5 * 10^{10} X_{t,3600}$	0.59643	[0.59640, 0.59647]

4.2 Zwei-Stichproben-t-Tests

Ein Zwei-Stichproben-t-Test prüft, ob sich die empirischen Mittelwerte zweier Stichproben bedeutend unterscheiden.

Der p-Wert ist in allen sechs Fällen geringer als das festgelegte Signifikanzniveau. Damit wird jeweils die Nullhypothese verworfen, dass die empirischen Mittelwerte gleich sind.

In den Testergebnissen in Tabelle 4.3 sind für die Unterschiede der empirischen Mittelwerte Konfidenzintervalle angegeben. Die Unterschiede der empirischen Mittelwerte werden für beide Lagemaße \overline{M}_n und \overline{M}_w im Zeitverlauf größer. Dieser Effekt konnte bisher noch nicht beobachtet werden.

Abschließend kann festgestellt werden, dass sich die empirischen Mittelwerte signifikant unterscheiden und die Ensemblegrößen einen signifikanten Einfluss auf das Simulationsergebnis haben. Bei der größeren Molekülanzahl kann eine höhere Präzision erwartet werden.

Tab. 4.3: Zwei-Stichproben-t-Tests für die beiden Lagemaße

Testgröße	p-Wert	emp. M.w. $1 * 10^{11}$	emp. M.w. $5 * 10^{10}$	95%-Differenz-Intervall
$\overline{M}_{n,600}$	1.198e-06	9933.57	9932.65	[0.55310, 1.29833]
$\overline{M}_{n,1800}$	9.146e-11	8540.30	8539.06	[0.87060, 1.62008]
$\overline{M}_{n,3600}$	<2.2e-16	7287.23	7284.24	[2.67084, 3.30577]
$\overline{M}_{w,600}$	3.57e-05	19801.72	19800.04	[0.88456, 2.47408]
$\overline{M}_{w,1800}$	1.348e-11	17333.93	17331.57	[1.68090, 3.04171]
$\overline{M}_{w,3600}$	<2.2e-16	15372.36	15367.80	[4.09089, 5.04099]

Anhand der Testergebnisse in Tabelle 4.4 ist ersichtlich, dass die Unterschiede der empirischen Mittelwerte für den Umsatz des Monomers X_t im Zeitverlauf etwas größer werden. Dieser Effekt konnte ebenfalls noch nicht beobachtet werden.

Tab. 4.4: Zwei-Stichproben-t-Tests für den Umsatz des Monomers

Testgröße	p-Wert	emp. M.w. $1 * 10^{11}$	emp. M.w. $5 * 10^{10}$	95%-Differenz-Intervall
$X_{t,600}$	0.0001182	0.18350	0.18362	[-0,0001939, -0,0000632]
$X_{t,1800}$	<2.2e-16	0.41356	0.41382	[-0.0003295, -0.0002075]
$X_{t,3600}$	<2.2e-16	0.59604	0.59644	[-0.0004513, -0.0003632]

4.3 Varianz-Tests

Ein Varianz-Test prüft, ob sich die theoretischen Varianzen zweier Stichproben bedeutend unterscheiden.

Der p-Wert ist in allen sechs Fällen geringer als das festgelegte Signifikanzniveau. Damit wird jeweils die Nullhypothese verworfen, dass die Varianzen gleich sind.

Es konnte bereits festgestellt werden, dass die empirischen Varianzen der Lagemaße für die größere Molekülanzahl deutlich kleiner sind. Aus den Ausgaben der Tests in Tabelle 4.5 kann nun herausgelesen werden, dass die Varianzen für die Lagemaße \overline{M}_n und \overline{M}_w bei Simulationen mit $5 * 10^{10}$ Molekülen in allen Fällen fast doppelt so groß sind wie bei $1 * 10^{11}$ Molekülen.

Diese Ergebnisse unterstreichen die Feststellung, dass sich die Ensemblegrößen signifikant auf das Ergebnis auswirken.

Tab. 4.5: Varianz-Tests für die beiden Lagemaße

Testgröße	p-Wert	Verhältnis emp. Varianzen	95%-Konfidenzintervall
$\overline{M}_{n,600}$	<2.2e-16	0.51364	[0.45370, 0.58151]
$\overline{M}_{n,1800}$	<2.2e-16	0.54832	[0.48433, 0.62076]
$\overline{M}_{n,3600}$	<2.2e-16	0.52234	[0.46138, 0.59135]
$\overline{M}_{w,600}$	<2.2e-16	0.50521	[0.44626, 0.57196]
$\overline{M}_{w,1800}$	<2.2e-16	0.58232	[0.51437, 0.65926]
$\overline{M}_{w,3600}$	<2.2e-16	0.57883	[0.51128, 0.65531]

Für den Umsatz des Monomers konnte bereits in Abschnitt 3.3 festgestellt werden, dass die Varianzen für die größere Molekülanzahl deutlich kleiner sind. Aus den Testergebnissen in Tabelle 4.6 lässt sich schließen, dass die Varianzen des Umsatzes des Monomers X_t für die kleinere Molekülanzahl von $5 * 10^{10}$ jeweils etwa doppelt so groß sind wie für eine Molekülanzahl von $1 * 10^{11}$.

Tab. 4.6: Varianz-Tests für den Umsatz des Monomers

Testgröße	p-Wert	Verhältnis emp. Varianzen	95%-Konfidenzintervall
$X_{t,600}$	<2.2e-16	0.51602	[0.45580, 0.58420]
$X_{t,1800}$	<2.2e-16	0.55107	[0.48676, 0.62388]
$X_{t,3600}$	<2.2e-16	0.50024	[0.44186, 0.56633]

5 Zusammenfassung

Es konnte anhand von Boxplots, Histogrammen und Konfidenzintervallen festgestellt werden, dass die Unterschiede in den Ensemblegrößen sehr klein sind (kleiner als eine Monomereinheit) und erst recht kleiner als im Vergleich zu experimentellen Ergebnissen, wo diese Präzision gar nicht erreicht wird.

Ferner zeigten die Zwei-Stichproben-t-Tests und die Varianz-Tests, dass die Ensemblegröße einen statistisch signifikanten Einfluss auf das Ergebnis hat, auch wenn dieser sehr klein ist.

Auf die Frage, ob das mehr an Informationstiefe einen erhöhten Berechnungsaufwand rechtfertigt, kann nun Stellung genommen werden. Natürlich geben Simulationsergebnisse mit einer höheren Molekülanzahl präzisere Ergebnisse. Allerdings haben die Simulationsergebnisse schon bei einer Molekülanzahl von $5 \cdot 10^{10}$ eine sehr viel höhere Präzision als experimentelle Ergebnisse und können als ausreichend bewertet werden.

A Anhang

A.1 Datenbasis

Auf dem beiliegenden Datenträger der Arbeit sind die Simulationsergebnisse für beide Molekülanzahlen sowie der Simulator als .zip-Datei gespeichert.

Es handelt sich um die Dateien *1e11.zip*, *5e10.zip* und *master.zip*.

A.2 Datenauswertung

Ferner finden sich auf dem Datenträger die vollständigen R-Skripte und .csv-Dateien der Datenauswertung.

Die Dateinamen lauten *CLD-Auswertung.R*, *D.600.0.cld.csv*, *D.3600.0.cld.csv* (im Unterordner 1CLD-Auswertung), *e10.MnMw.csv*, *e11.MnMw.csv*, *MnMw-Auswertung.R* (im Unterordner 2MnMw-Auswertung) sowie *e10.c.csv*, *e11.c.csv* und *Umsatz-Auswertung.R* (im Unterordner 3Umsatz-Auswertung).

A.3 Dokumentation

Schließlich ist eine digitale Ausfertigung der Arbeit im .pdf-Format auf dem Datenträger gespeichert.

Die Datei trägt den Namen *Bachelorarbeit.pdf*.

Literaturverzeichnis

- [1] B. Tieke, *Makromolekulare Chemie*, 3. Aufl., Wiley-VCH, **2014**.
- [2] M. Drache et al., *Macromolecular Theory and Simulations* **2015**, 24, 301–310.
- [3] P. W. Atkins, J. de Paula, *Physikalische Chemie*, 5. Aufl., Wiley-VCH, **2013**.
- [4] M. D. Lechner et al., *Makromolekulare Chemie*, 5. Aufl., Springer Spektrum, **2014**.
- [5] D. J. Hart et al., *Organic Chemistry: A Short Course*, 13. Aufl., Brooks/Cole, **2012**.
- [6] G. Schwedt et al., *Analytische Chemie*, 3. Aufl., Wiley-VCH, **2016**.
- [7] L. Fahrmeir et al., *Statistik*, 8. Aufl., Springer Spektrum, **2016**.
- [8] D. T. Gillespie, *The Journal of Physical Chemistry* **1977**, 81, 2340–2361.
- [9] M. Drache, G. Drache, *Polymers* **2012**, 4, 1416–1442.
- [10] A. L. T. Brandão et al., *Macromolecular Reaction Engineering* **2015**, 9, 141–185.
- [11] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, **2018**.
- [12] P. Dalgaard, *Introductory Statistics with R*, 2. Aufl., Springer, **2008**.
- [13] J. Groß, *Grundlegende Statistik mit R*, 1. Aufl., Vieweg+Teubner, **2010**.
- [14] W. N. Venables, B. D. Ripley, *Modern Applied Statistics with S*, 4. Aufl., Springer, **2002**.

Hiermit versichere ich, dass ich die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und dass alle Stellen dieser Arbeit, die wörtlich oder sinngemäß aus anderen Quellen übernommen wurden, als solche kenntlich gemacht wurden und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsstelle vorgelegt wurde.

Des Weiteren erkläre ich, dass ich mit der öffentlichen Bereitstellung meiner Abschlussarbeit in der Instituts- und Universitätsbibliothek einverstanden bin.

Clausthal-Zellerfeld, 30. Januar 2019